

NMPDR Update

Rick Stevens

University of Chicago

Argonne National Laboratory

The Fellowship for Interpretation of Genomes

University of Illinois Urbana Champaign



www.nmpdr.org



Some Highlights Since the Last BRC Meeting

- Inclusion of essentially all complete genomes in NMPDR to support comparative analysis
- Acceleration of our update and data release cycle to approximately two weeks (version 5)
- Integration of BRC data via BRC central as external expert sources in NMPDR
- Beginning collection of annotation statistics
- Subsystems extension beyond conserved core
- Started development of an in silico drug target development pipeline
- Multiple workshops, posters and tutorials

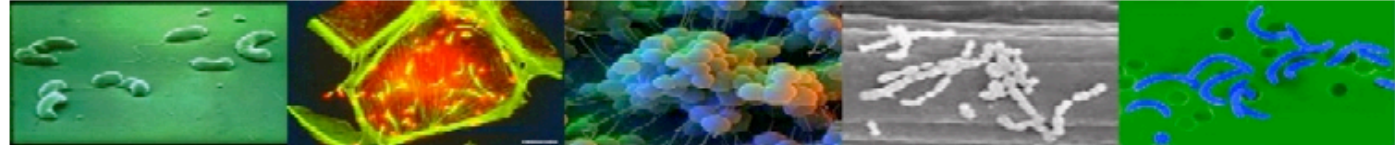


Current NMPDR Usage Statistics

- About 1,000 requests per day
- Downloading about ~50 MB per day
- 35% .edu, 33% .com, 14% .net, 4% .gov, 2% .ca (Canada), > 1% (France, Germany, Thailand, Brazil, Japan, Norway, Mexico, UK, Malaysia, Australia, Belgium, Italy, Sweden)
- Significant number coming in via organism search terms (genus mainly)



NMPDR

[Home](#)[About Us](#)[Using the NMPDR \(FAQs\)](#)[Events](#)[News](#)[Links](#)[▶ Links to Other Resources](#)[▶ Links to Other BRC Sites](#)[▶ Links to RCE Sites](#)[Organism Summaries](#)[▶ Campylobacter](#)[▶ Listeria](#)[▶ Staphylococcus](#)[▶ Streptococcus](#)[▶ Vibrio](#)[Organism Data](#)[▶ Search](#)[▶ Downloads](#)[Subsystem Summaries](#)[NMPDR Tools](#)[Contact Us](#)[Mailing Lists](#)[Sitemap](#)[How to Cite the NMPDR Web Pages](#)[NMPDR Database Documentation](#)

National Microbial Pathogen Data Resource

Version 5 (2/4/06)

includes 343 bacterial, 26 archaeal and 29 eukaryal genomes

NMPDR is a [Bioinformatics Resource Center](#) dedicated to the thorough understanding of [core organisms](#) assigned by the [National Institute of Allergy and Infectious Diseases](#). NMPDR is an environment for the comparative functional analysis of genomes.

Statistics describing each of the NMPDR core organisms, including the number of both proven and hypothetical genes included in subsystems, may be found in this table of [NMPDR Genome Statistics](#).

Search NMPDR Database:

Search

Advanced

keywords or accession number

Begin by entering **keywords** in the box above to **search** for a gene or protein to work on. Keywords may be the names of genes, proteins, or organisms. Keywords may also be alphanumeric labels such as Enzyme Commission (EC) numbers or accession numbers assigned by other data resources. The search joins multiple terms with "AND," so will return a table listing records that match all words in the request.

Another way to begin is to go to an **organism summary** page (linked at left) and do a keyword search for your protein of interest from there. This will automatically limit the search to your focus group of organisms. You may also select a genome or organism to search by clicking the **Advanced** button above.

To search using the sequence of a protein or gene as your query, please click the **Advanced** button to use **BLAST**. You must choose which organisms to BLAST against from the drop-down list. For more instruction, please see [Searching NMPDR](#).

Continue from the table of text search results, which will allow you to examine your protein in the GBrowse or NMPDR environment. Click the NMPDR button to view **your protein (highlighted green)** in the context of its neighbors in close proximity on the genome. Clicking the Pins button will open a graphic display of homologous regions in many genomes. Clicking the CL button will return a table listing homologs that cluster with functionally related neighbors in other genomes. For more instruction, please see the list of frequently asked questions, [Using the NMPDR](#), or the tutorial, [Navigating NMPDR](#).

Integrating BRC Central Data Sources into NMPDR

The NMPDR Protein Page

http://www.nmpdr.org/FIG/protein.cgi?prot=fig|273123.1.peg.3414&user=&SPROUT=1

KEGG DAS Database distribution Automator Pr...ipt Actions Mayberrys ClieSource MCS Mac Software Google http://yersi...bDetails.cgi Radio-Locator ZINC is not ...al screening tipizen Compulab Drug >>

NMPDR National Microbial Pathogen Data Resource Center

National Microbial Pathogen Database Resource

[Search the NMPDR: | Submit a bug report]

[Pathema | VectorBase | VBRC | ERIC | BioHealthBase | ApiDB | PATRIC]
 [GOLD | NCBI | KEGG | ExPASy | TIGR cmr | UniProt | MetaCyc]
 [MBGD | Milton Sailer's Transport Links | Ian Paulsen's TransportDB | 123Genomics | Microbes.info | The Enzyme List | Enzyme Corner]

Protein fig|273123.1.peg.3414: Yersinia pseudotuberculosis IP 32953 [273123.1]

[Translate Function Assignments](#)

Context on contig 273123.1:NC_006155 from base 3968004 to 3978872 (10869 bp)

fid	starts	ends	size	gap	find best clusters	pins	fc-sc	SS	Ev	comment	aliases
3408	3969575	3968790	786	-	CL			1		Flagellar basal-body rod protein flgG	YPTB3335, YP_071826.1, eric ABI-0033223, eric ABY-0117710, eric ACQ-0241590, eric ADI-0000530, flgG, gil51597635, kegglype:YPO0728, kegglypk:y3450, kegglypm:YP3039, kegglyps:YPTB3335, trjQ666C3, trjQ8ZH26, unilQ666C3, unilQ8ZH26
3409	3970374	3969643	732	-	CL			1		Flagellar basal-body rod protein flgF	YPTB3336, YP_071827.1, flgF, gil51597636, kegglyps:YPTB3336, trjQ666C2, unilQ666C2
3410	3971618	3970374	1245	-	CL			1		Flagellar hook protein flgE	YPTB3337, YP_071828.1, flgE, gil51597637, kegglyps:YPTB3337, trjQ666C1, unilQ666C1
3411	3972329	3971673	657	-	CL	Pins	4	1		Flagellar basal-body rod modification protein flgD	YPTB3338, YP_071829.1, flgD, gil51597638, kegglyps:YPTB3338, trjQ666C0, unilQ666C0
3412	3972757	3972332	426	-	CL			2		Flagellar basal-body rod protein flgC	YPTB3339, YP_071830.1, eric ABI-0033228, eric ABY-0117700, eric ACQ-0241586, eric ADI-0000526, flgC, gil51597639, kegglype:YPO0723, kegglypk:y3455, kegglypm:YP3035, kegglyps:YPTB3339, trjQ666B9, trjQ8ZH29, unilQ666B9, unilQ8ZH29
3413	3973206	3972760	447	-	CL			2		Flagellar basal-body rod protein flgB	YPTB3340, YP_071831.1, eric ABI-0033229, eric ABY-0117697, eric ACQ-0241582, flgB, gil51597640, kegglype:YPO0722, kegglypk:y3456, kegglypm:YP3034, kegglyps:YPTB3340, trjQ666B8, trjQ8ZI00, unilQ666B8, unilQ8ZI00
3414	3973294	3974112	819	+	CL	Pins		1		Flagellar basal-body P-ring formation protein flgA	YPTB3341, YP_071832.1, flgA, gil51597641, kegglyps:YPTB3341, trjQ666B7, unilQ666B7
3415	3974322	3974588	267	+	CL					Putative flagellar regulatory protein	YPTB3342, YP_071833.1, eric ABI-0033231, eric ABY-0117695, eric ACQ-0241580, eric ADI-0000523, gil51597642, kegglype:YPO0720, kegglypk:y3458, kegglypm:YP3032, kegglyps:YPTB3342, trjQ666B6, trjQ8ZI02, unilQ666B6, unilQ8ZI02
3416	3974595	3975020	426	+	CL					Hypothetical protein	YPTB3343, YP_071834.1, gil51597643, kegglyps:YPTB3343, trjQ666B5, unilQ666B5
3417	3975799	3975362	438	-	CL					Hypothetical	YPTB3344, YP_071835.1, gil51597644, kegglyps:YPTB3344, trjQ666B4, unilQ666B4



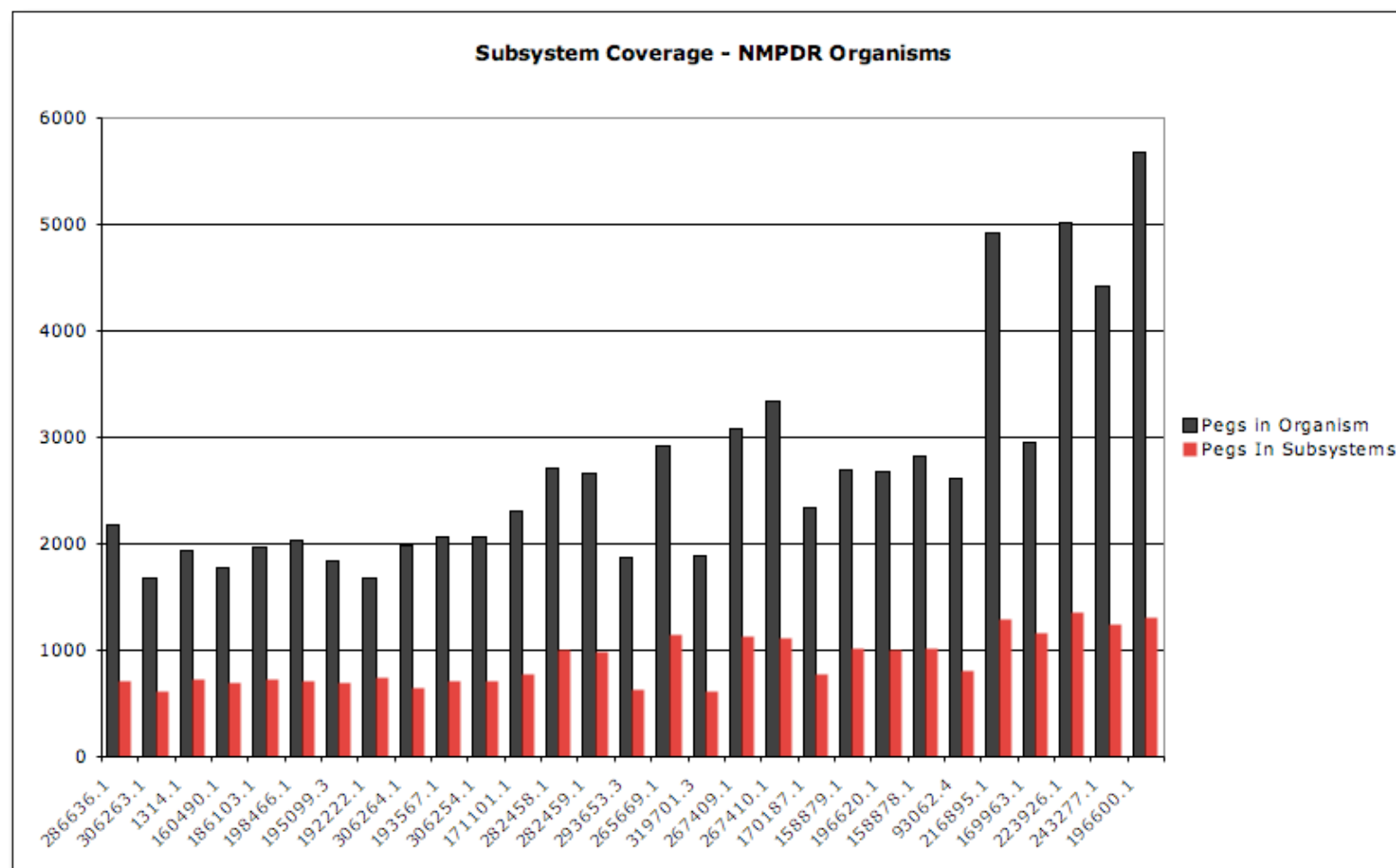
www.nmpdr.org

NMPDR



National Microbial
Pathogen Data
Resource Center

Subsystem Coverage - NMPDR



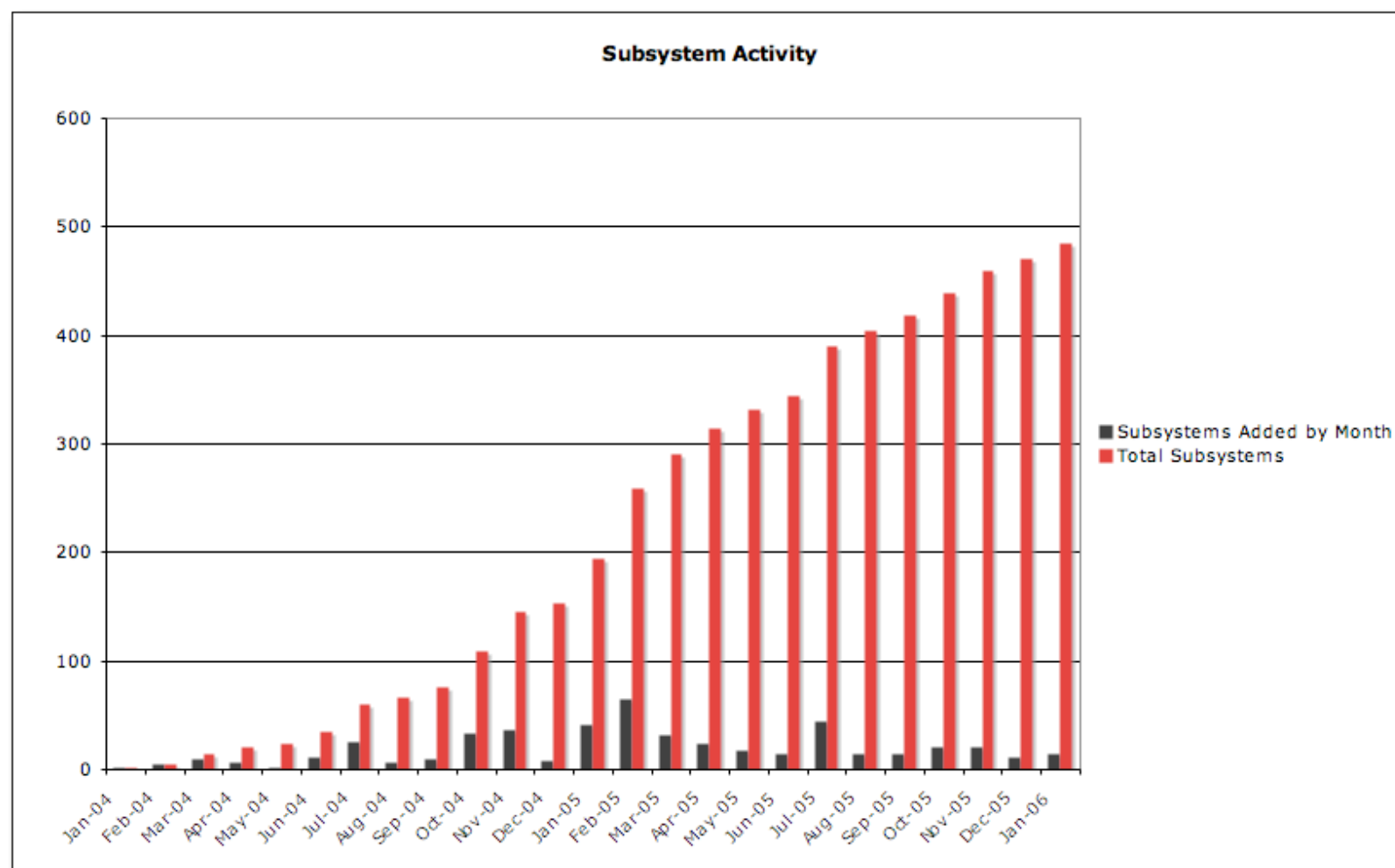
www.nmpdr.org

NMPDR



National Microbial
Pathogen Data
Resource Center

Subsystems Activity



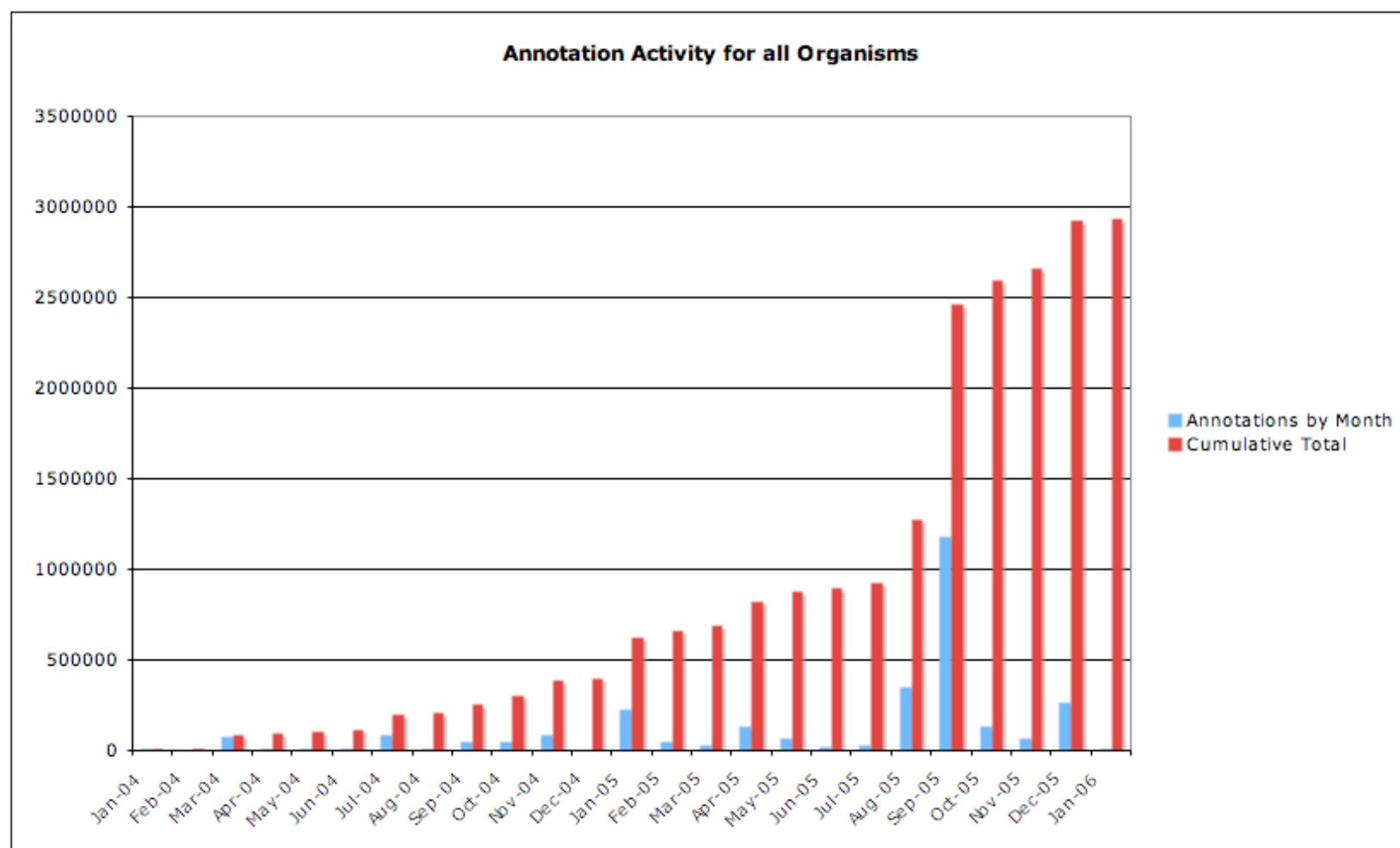
www.nmpdr.org

NMPDR



National Microbial
Pathogen Data
Resource Center

Annotation Activity All Organisms



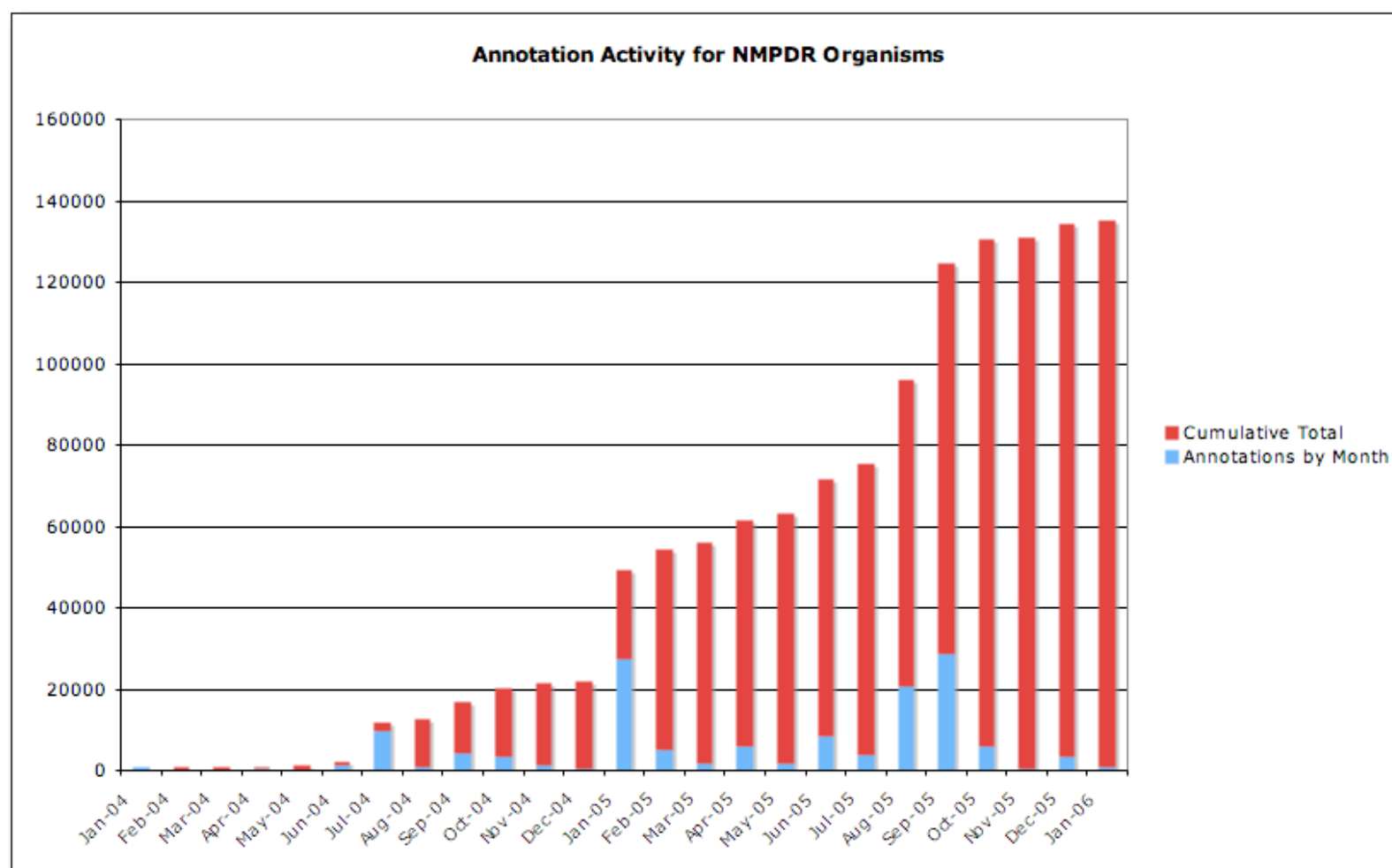
www.nmpdr.org

NMPDR



National Microbial
Pathogen Data
Resource Center

Annotation Activity for NMPDR



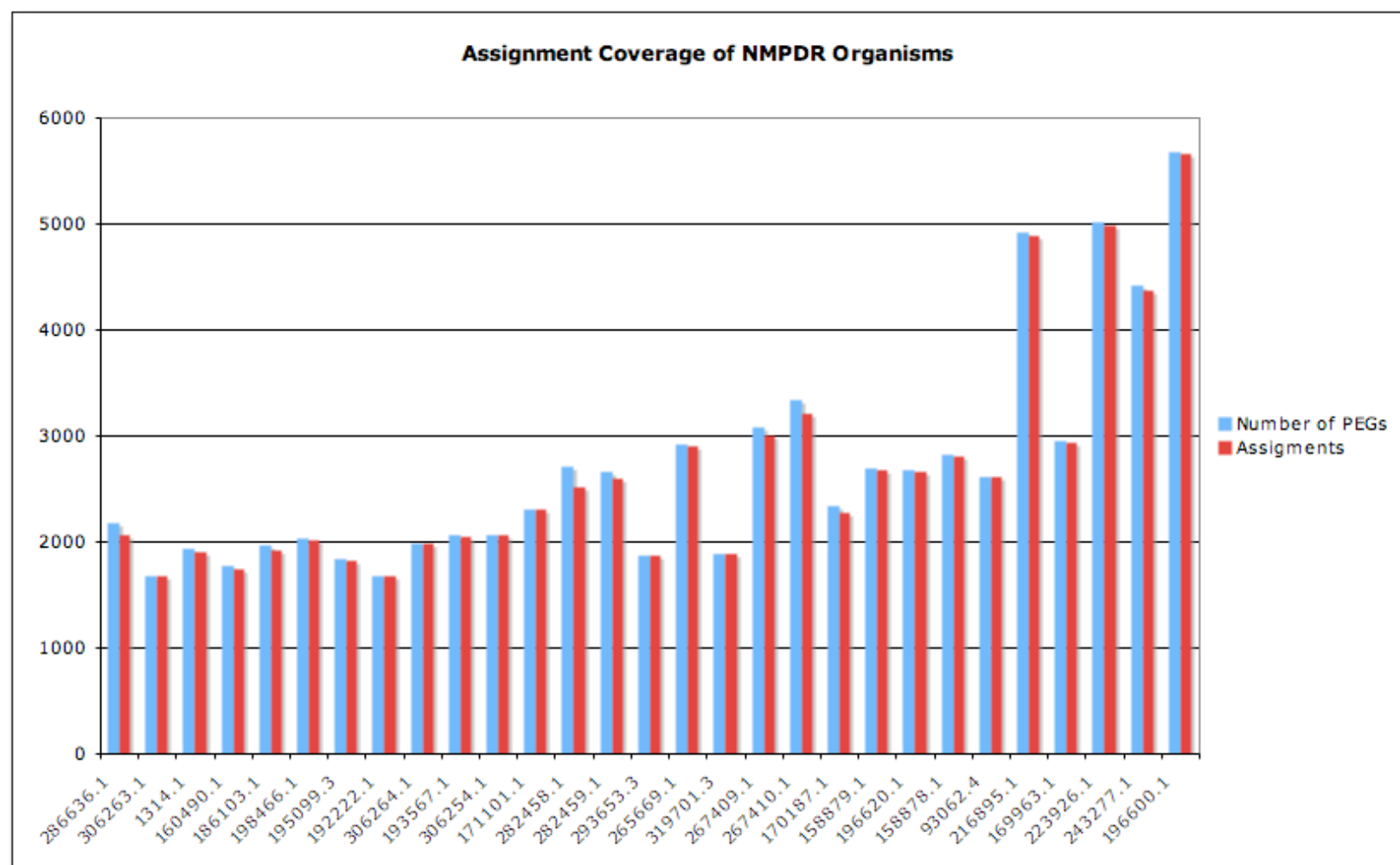
www.nmpdr.org

NMPDR



National Microbial
Pathogen Data
Resource Center

Assignment Coverage for NPMDR



www.nmpdr.org

NMPDR



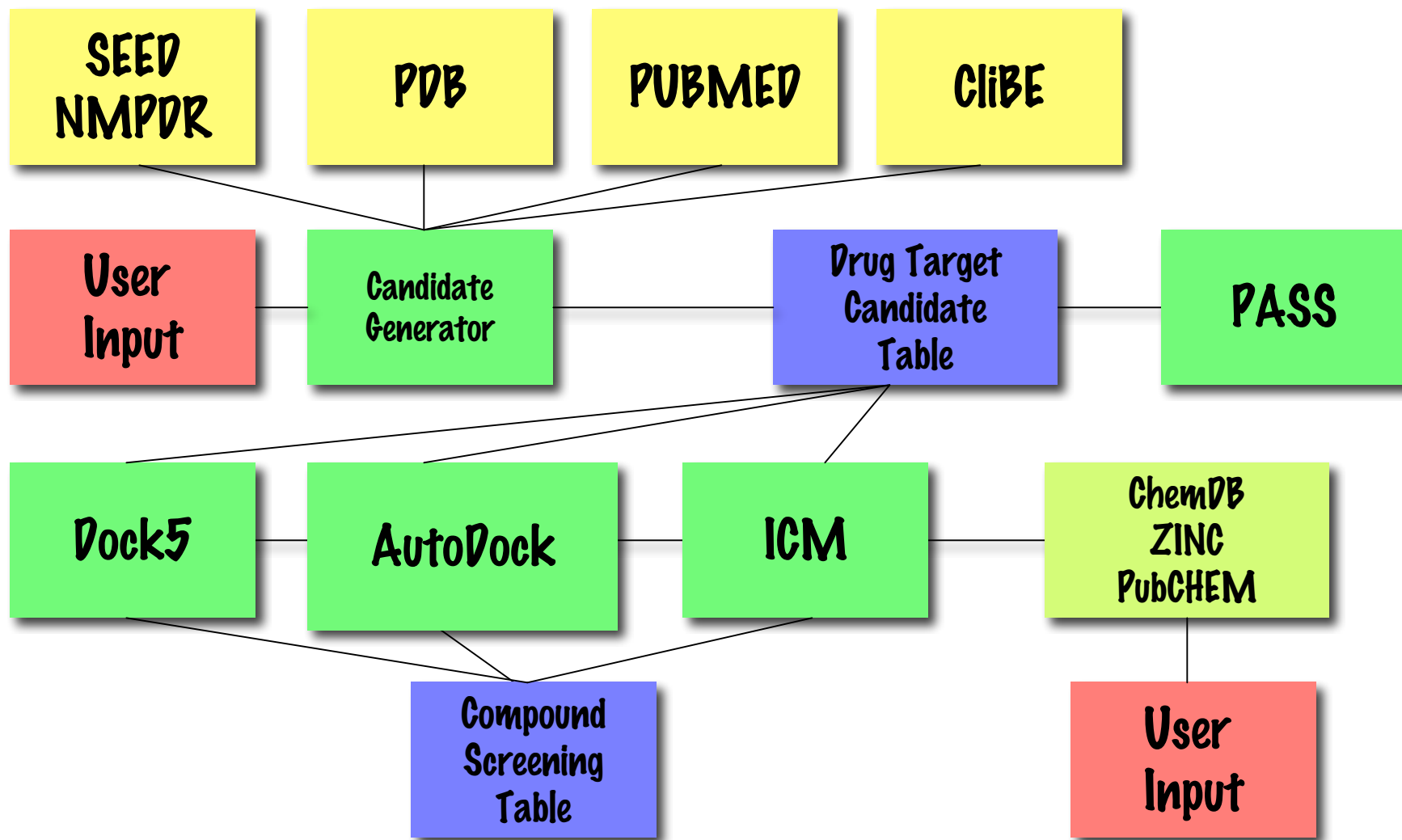
National Microbial
Pathogen Data
Resource Center

Virulence-related Subsystems in NMPDR

- **Attachment and colonization factors;**
 - *Pili and fimbriae*
 - *Nonfimbriar adhesins*
- **Multiplication and Nutrition in host:**
 - *Iron scavenging mechanisms*
- **Invasion and intracellular survival**
- **Evading host defense mechanisms**
 - *Capsules, extracellular polymers*
 - *Antigenic variation*
- **Antibiotic resistance**
- **Toxins**
- **Pathogen. islands, prophages, gene clusters**
- **Other virulence factors:**
 - *Regulation of virulence*
 - *Secretion systems*
 - *Motility and hemotaxis*
 - *Classification of virulent strains*
- **Entire Virulome**

<i>Staphylo-ccocus</i>	<i>Strepto-ccocus</i>	<i>Listeria</i>	<i>Campylo-bacter</i>	<i>Vibrio</i>
				●
	●	●	●	
●	●	●		
	●	●		
	●			●
	●	●		
●●●●	●●●●	●●	●●	●
	●●		●	●
	●	●		●
●	●●●●	●	●●	●
●	●	●		
●	●	●	●	●
	●			

Drug Target Development and Screening



Future Plans

- Comprehensive “virulome” annotation
- Microarray data integration with subsystems
- Inference and computed evidence for assignments
- GO terms integration via subsystems and UniProt
- Reactions assigned to functional roles to support metabolic modeling
- Integration of promoters, transcription factors and regulatory sites with subsystems
- Detailed gene family histories and evidence for horizontal gene transfer
- Rapid assignment propagation capability for new genomes

